Accumulation and functional architecture of deleterious genetic variants during the extinction of Wrangel Island Mammoths

Erin Fry[1+], Sun K. Kim[2+], Sravanthi Chigurapti[1+], Katelyn M. Mika[1], Aakrosh Ratan[3], Alexander Dammermann[4], Brian J. Mitchell[2], Webb Miller[5], Vincent J. Lynch[1*]

[1] Department of Human Genetics, The University of Chicago, 920 E. 58th Street, CLSC 319C, Chicago, IL 60637, USA.

[2] Department of Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA.

[3] Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, 22908, USA

[4] Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9, A-1030 Vienna, Austria.

[5] Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, 506B Wartik Lab, University Park, PA 16802, USA.

[+]These authors contributed equally

[*]Correspondence: vjlynch@uchicago.edu

Woolly mammoths were among the most abundant cold adapted species during the Pleistocene. Their once large populations went extinct in two waves, an end-Pleistocene extinction of continental populations[1-4] followed by the mid-Holocene extinction of relict populations on St. Paul Island ~5,600 years ago[5] and Wrangel Island ~4,000 years ago[1-4]. Population and conservation genetics theory predicts that deleterious alleles will accumulate as populations decline, leading to downward spiral of declining population size and fitness ending in extinction (mutational meltdown). While the extinction of Wrangel Island mammoths was preceded by prolonged demographic decline, reduced population size and genetic diversity, recurrent inbreeding, and the fixation of deleterious alleles[3,4,6-8], the functional consequences of these processes are unclear. Here we show that the extinction of Wrangel Island mammoths was accompanied by an accumulation of deleterious mutations that are predicted to cause diverse behavioral and developmental defects. Resurrecting and functional characterization of Wrangel Island mammoth genes with putative deleterious substitutions identified both loss and gain of function mutations associated with ciliopathies, oligozoospermia and reduced fertility, and neonatal diabetes. These results indicate that last mammoths likely suffered from genetic disease that reduced fitness and directly contributed to their extinction.

Woolly mammoths (*Mammuthus primigenius*) were among the most abundant cold adapted megafaunal species during the Middle to Late Pleistocene (*ca* 116-12 Kyr BP), inhabiting a large swath of steppe-tundra that extended from Western Europe, through Asia and Beringia, into North America. Paleontological and genetic data indicate mammoths went extinct in at least two waves, an end-Pleistocene decline and extinction of continental populations[1-4] followed by the mid-Holocene extinction of relict populations on St. Paul Island ~5,600 years ago[5] and Wrangel Island ~4,000 years ago[1-4] (**Fig. 1**). Extinction of Wrangel Island mammoths was preceded by prolonged demographic decline, resulting in a small population, reduced genetic diversity, recurrent breeding among distant relatives, and the fixation of deleterious alleles[3,4,6-8]. The functional significance of putatively deleterious amino acid substitutions in the Wrangel Island mammoth, however, is unknown. Thus it is unclear if an accumulation of deleterious variants contributed to the extinction of Wrangel Island mammoths.

To infer if the extinction of Woolly mammoths was associated with an accumulation of putatively deleterious amino acid variants, we identified homozygous nonsynonymous substitutions unique (private) in three extant Asian elephants (*Elephas maximus*) and three woolly mammoths, the ~44,800 year old Oimyakon mammoth [1], the ~20,000 year old M4 mammoth [9-13], and the ~4,300 years old Wrangel Island mammoth [1]. These mammoths span the age from when mammoth populations were large and wide-spread (Oimyakon), to near the beginning of their final decline (M4), and their last known population (Wrangel Island); thus these individuals allow us to infer if deleterious mutations accumulated in the mammoth genome during their decline and eventual extinction. Next we used PolyPhen-2 [14,15] to computationally predict the functional impact of each private homozygous amino acid substitution (**Table 1**).

The proportion of 'possibly damaging' amino acid substitutions in the Wrangel Island genome (109/573=0.190) was significantly greater (binomial test N≥109, P=0.011) than Oimyakon and M4 (40/259=0.154), whereas the proportion of 'benign' (binomial test N≥349, P=0.979) and 'probably damaging' (binomial test N≥115, P=0.429) amino acid substitutions in the Wrangel Island genome was not significantly different than Oimyakon and M4 (**Fig. 2a**). Genes with homozygous 'probably damaging' amino acid variants in the Wrangel Island mammoth were enriched for 22 mouse knockout phenotypes at an FDR≤0.25 (**Supplementary Table 1 and 2**), four human phenotypes at an FDR≤0.47 (**Supplementary Table 3**), and nine disease ontologies at an FDR≤0.71 (**Supplementary Table 4 and 5**) compared to M4 and Oimyakon (**Fig. 2b**). These results suggest that there was an increased burden of weakly deleterious amino acid substitutions in the Wrangel Island mammoth.

Genes with 'probably damaging' amino acid variants in the Wrangel Island mammoth were enriched for numerous mouse KO phenotypes related to behavioral and neurological defects such as 'abnormal postural reflex' (E=7.65, hypergeometric P=0.018, FDR q=0.25), 'altered righting response' (E=4.59, hypergeometric P=0.017, FDR q=0.25), 'abnormal posture' (E=7.65, hypergeometric P=0.017, FDR q=0.25), and 'abnormal maternal nurturing' (E=7.65, hypergeometric P=0.017, FDR q=0.25), specifically pup cannibalization and tendency for females to trample pups. Among the most notable of these genes is hydrolethalus syndrome protein 1 (HYLS1), a centriolar protein that functions in ciliogenesis (Dammermann et al., 2009). Mutations in HYLS1 underlie hydrolethalus syndrome (HLS [MIM: 236680]), a perinatal lethal developmental disorder characterized by severe brain malformation including hydrocephalus

and absent midline structures (Mee et al., 2005), as well as Joubert syndrome (JBTS [MIM: 213300]), a milder disorder characterized by defects in the cerebellum and brain stem leading to impaired balance and coordination (Oka et al., 2016).

The Wrangel Island mammoth P119L mutation occurs in a highly conserved region of the protein, which is invariant for proline or serine across vertebrates and is therefore potentially deleterious (**Fig. 3a**). To infer if this mutation had functional consequences, we used a *Xenopus* model of ciliogenesis[16]. Morpholino oligos (MO) targeting *Xenopus HYLS1* led to a severe defect in cilia assembly (Wilcox test, P=9.48×10^{-6}; **Fig. 3c/h**), as previously reported [16]. This defect was rescued by addition of MO-resistant wild-type *Xenopus HYLS1* (Wilcox test, P=1.71×10^{-4}; **Fig. 3d/h**), but not a variant incorporating the equivalent P119L mutation into *Xenopus* HYLS1 (S186L) (Wilcox test, P=2.56×10^{-5}; **Fig. 3e/h**). The S186L mutant HYLS1 did, however, appropriately localize to centrioles and did not have any dominant negative effects in the absence of depletion of the endogenous protein (**Fig. 3f/g** and not shown). Given the widespread role of cilia in vertebrates, this mutation may have caused ciliopathy-like phenotypes including neurological and skeletal abnormalities and male infertility[17] in Wrangel Island mammoths.

Another gene with a 'probably damaging' amino acid substitution in the Wrangel Island mammoth associated with male infertility is *Naked cuticle 1* (*NKD1*), which encodes passive antagonist of the Wnt/TCF-LEF signaling pathway [18-20]. The Wrangel Island mammoth A88V substitution occurred at a site that is nearly invariant for alanine or serine in diverse vertebrates (**Fig. 4a**), suggesting it may have had functional consequences. To infer if the A88V substitution had functional affects, we resurrected the Wrangel Island/M4 ancestral (AncYakut, **Fig. 2a**) and Wrangel Island *NKD1* genes and tested their ability to antagonize Wnt-signaling in elephant dermal fibroblasts transiently transfected with a luciferase reporter vector containing a minimal promoter and eight copies of a TCF-LEF response element (pGL4.49[*luc2P*/TCF-LEF/Hygro]) and treated with a small molecule agonist of the Wnt-signaling pathway (CHIR99021). The AncYakut NKD1 reduced luminescence to background levels in response to CHIR99021 treatment (Wilcox test, P=2.71×10^{-6}), in stark contrast however, the Wrangel Island NKD1 did not affect luciferase expression (Wilcox P=0.98), indicating the A88V substitution is a loss of function mutation (**Fig. 4b**). Transgenic mice with loss of function mutations in NKD1 have dysregulated Wnt/beta-catenin signaling in the testis leading to abnormal seminiferous tubule

morphology, small seminiferous tubules, small testis, oligozoospermia, and reduced fertility [21,22], suggesting the A88V substitution may have affected fertility in Wrangel Island mammoths.

Although Wrangel Island mammoths were generally smaller than mainland mammoth populations, their size varied and large and small individuals coexisted on the Island. Four genes with 'probably damaging' amino acid substitutions in the Wrangel Island genome caused 'decreased body size' (MP:0001265) and five genes with 'probably damaging' amino acid substitutions caused 'decreased body weight' (MP:0001262) in mouse knockouts. In stark contrast, genes with 'probably damaging' amino acid substitutions were not associated with mouse knockout phenotypes such as 'increased body size' (MP:0001264) and only a single gene was associated with 'increased body weight' (MP:0001262). The proportion of genes with 'probably damaging' substitutions associated with 'decreased body size' and 'decreased body weight' phenotypes (7/38 = 0.184) was significantly greater (binomial test N≥7, P=4.48×10^{-5}) than the proportion of genes associated with 'increased body size' and 'increased body weight' phenotypes (1/38 = 0.026), suggesting these variants may have contributed to variation in size among Wrangel Island mammoths.

Among the genes with 'probably damaging' mutations in the Wrangel Island mammoth associated with decreased body size was *NEUROGENIN 3* (*NEUROG3*), which encodes a basic helix-loop-helix transcription factor that is required for endocrine cell development in the pancreas and intestine. The 'probably damaging' G195E substitution in the Wrangel Island mammoth NEUROG3 protein occurred at a site that is nearly invariant for glycine across mammals (**Fig. 5a/b**) within an LXXLL motif in the C-terminal transcriptional activation domain [23]. To infer if the G195E substitution had functional affects, we resurrected the AncYakut and Wrangel Island *NEUROG3* genes and tested their ability to transactivate luciferase expression from a reporter vector containing a minimal promoter and six repeats of the *PAX4* E-box (pGL3[*luc/6x-PAX4E/minP*]) in transiently transfected elephant dermal fibroblasts (**Fig. 5c**). The Wrangel Island NEUROG3 transactivated luciferase expression from the pGL3[*luc/6x-PAX4E/minP*] reporter vector more strongly than the AncYakut NEUROG3 protein (1.8-fold, Wilcox P=0.007), indicating the G195E substitution is a hypermorphic mutation. Loss of function mutations in the human *NEUROG3* gene cause congenital malabsorptive diarrhea (DIAR4 [MIM:610370]), a disorder characterized by neonatal diabetes, chronic unremitting malabsorptive diarrhea, vomiting, dehydration, and severe hyperchloremic metabolic acidosis [24-

[26]. *NEUROG3* knock-out mice die postnatally from diabetes[26] suggesting the G195E substitution may have affected insulin signaling in Wrangel Island Mammoths.

Previous studies of Wrangel Island mammoths have found signatures of reduced genetic diversity, reduced population size, and recurrent breeding among distant relatives [1], and an increased burden of deletions, retrogenes, genes with deleted exons, and genes with premature stop codons [6]. The functional and phenotypic consequences of these findings, however, are unclear. We found that the Wrangel Island mammoth genome had significantly more putatively deleterious mutations than older continental mammoths, and that these mutations were associated abnormal phenotypes in knockout mice. These data suggest that Wrangel Island mammoths suffered from genetic disorders resulting from the accumulation of deleterious mutations in an inbred and rapidly declining population. Indeed, our functional assays suggest that the Wrangel Island mammoth suffered from ciliopathies, reduced male fertility, and pathological insulin signaling. Thus an accumulating burden of deleterious mutations in Wrangel Island mammoths may have contributed to their extinction.

METHODS

Genome assembly

Details of the sequencing protocol for the Oimyakon and Wrangel Island mammoths can be found in Palkopoulou et al. (2015) and for the Asian elephants, M25, and M4 in Lynch et al. (2015). Sequences from the Asian elephants were aligned to the reference genome from the African Elephant (loxAfr3) using the Burrows Wheeler Aligner (Li and Durbin, 2010) with default parameters (BWA version 0.5.9-r16). The reads were subsequently realigned around putative indels using the GATK (DePristo et al., 2011) IndelRealigner (version 1.5-21-g979a84a), and putative PCR duplicates were flagged using the MarkDuplicates tool from the Picard suite (version 1.96).

The sequences from the mammoths were treated separately to account for DNA damage in the sequences. Putative adapter sequences were removed and we merged overlapping paired-end reads using available scripts (Kircher, 2012). We required an overlap of at least 11 nucleotides between the mates, and only pairs that could be merged were retained for subsequent analyses. The merged reads were aligned to the genome from the African elephant (loxAfr3) using BWA with default parameters, and only the mapped reads that were longer than 20 bps were retained for the subsequent SNP calls. The reads were realigned using the GATK IndelRealigner and putative PCR duplicates were flagged using MarkDuplicates, similar to the process described for the modern genomes. We also limited the incorporation of damaged sites into the variant-calling pipeline by hard-masking all sites that would be potentially affected by the characteristic ancient DNA patterns of cytosine deamination in single stranded overhangs. This mask was applied to 10 nucleotides on both ends of the merged reads from the ancient samples.

Single-nucleotide variants (SNVs), i.e. positions in the African elephant reference assembly at which we detected a nucleotide different from the reference in at least one of the Asian elephant or mammoth individuals were identified using SAMtools (Li et al., 2009) (version 0.1.19), which was applied with “-C50” to adjust the mapping quality of the reads with multiple mismatches. We did not call differences in regions where the reference base was unknown, and the calls were limited to regions that were covered at least 4 times, and at most 350 times by the sequences in these samples. We then identified homozygous nonsynonymous substitutions

unique (private) to each elephant and mammoth genome; these homozygous nonsynonymous substitutions were used in downstream analyses.

The M25 genome is likely a composite of multiple mammoth individuals [6], therefore, we did not include M25 in downstream analyses. However, because we are primarily interested in private nonsynonymous variants within each elephant and mammoth genome we were able advantage of the composite nature of the M25 genome by excluding homozygous nonsynonymous substitutions identified in the three Asian elephants and the three mammoths that were also observed in M25. The rationale for this filtering process is that any homozygous nonsynonymous substitution observed in either the Asian elephants or the three mammoths that is also observed in M25 is not truly a private variant. We thus identified 106 fixed amino acid substitutions in 99 genes in the Oimyakon mammoth, 162 fixed amino acid substitutions in 143 genes in the M4 mammoth, and 594 fixed amino acid substitutions in 525 genes in the Wrangel Island mammoth.

Functional annotation

We used PolyPhen-2 to classify amino acid substitutions as 'benign', 'possibly damaging', or 'probably damaging' [14,15] and GREAT [27] to infer the functional consequences of fixed 'damaging' amino acid substitutions in each mammoth. While GREAT was initially designed to annotate non-coding regions (cis-regulatory elements in particular), slight modifications allow similar functional annotation of coding regions (http://bejerano.stanford.edu/forum/viewtopic.php?f=2&t=990). Specifically, we mapped elephant genes to the to the transcription start site of the canonical transcript of their homologs in the UCSC 'knownCanonical' gene set and used the 'Single nearest gene' association rule in GREAT. The foreground dataset for GREAT included Wrangel Island either all genes with 'probably damaging' or all genes with 'possibly damaging' and 'probably damaging' substitutions, the background dataset for both foreground genes sets were all genes with private homozygous nonsynonymous substitutions in the Oimyakon, M4, and Wrangel Island genomes.

Data availability

Homozygous nonsynonymous substitutions unique (private) in the three extant Asian elephants and mammoths and PolyPhen-2 functional annotations for each variant are available at Dryad.

HYLS1 functional validation

Xenopus embryos were acquired by in vitro fertilization using standard protocols (Sive et al., 2000) approved by the Northwestern University Institutional Animal Care and User Committee. Previously validated MOs (GeneTools) were used (Control MO, 5'-CCTCTTACCTCAGTTACAATTTATA-3'; HYLS-1.1, 5'-GAACTGCCTGTCTCGAAGTGACATG-3'; XHYLS-1.2, 5'-GAACTGCCTGTCTCTCAGTGACATG-3' (Dammermann 2009). Full length XHYLS1 and the Wrangel mammoth mutant *Xenopus* equivalent XHYLS1-S186L were cloned into pCS2+ and fused with GFP at the N terminus. mRNA of the pCS2 constructs was prepared using in vitro transcription with SP6 (Promega). Morpholinos and mRNA were coinjected into each blastomere at the 2-4 cell stage using a total of 50-75 ng of morpholino and 500pg-1ng mRNA per embryo. Embryos were allowed to develop until stage 28 then fixed with 4%PFA in PBS for 2 hrs at RT. For antibody staining embryos were blocked for 1 hr in PBS with 0.1% Triton and 10% Normal Goat Serum prior to overnight incubation with primary antibody (Acetylated tubulin, Sigma T6793). Fluorescent secondary Abs (Jackson Labs) were incubated overnight after a full day of washing in PBS-0.1 %Triton. After secondary washing embryos were stained with fluorescently tagged phalloidin to mark the cell boundaries. Imaging was performed on a laser-scanning confocal microscope (A1R; Nikon) using a 60× oil Plan-Apo objective with a 1.4 NA.

NKD1 functional validation

To infer if the A88V substitution had functional affects, we synthesized the ancestral mammoth (AncYakut, A88) and Wrangel Island (V88) *NKD1* genes with mouse codon usage and cloned each gene into the mammalian expression vector pcDNA3.1+C-DYK. Next we tested their ability to antagonize luciferase expression from the pGL4.49[*luc2P*/TCF-LEF/Hygro luciferase reporter vector, which drives luciferase expression from a minimal promoter and eight copies of a TCF-LEF response element upon activation of Wnt-signaling. African elephant primary dermal fibroblasts (San Diego Zoo, "Frozen Zoo") were grown at 37°C/5% CO2 in a culture medium consisting of FGM/MEM (1:1) supplemented with 10% FBS, 1% Glutamine, and 1% penstrep. Confluent cells in 96-well plates in 60 µL of Opti-MEM (GIBCO) were transfected with 100 ng of the luciferase reporter plasmid pGL4.49[*luc2P*/TCF-LEF/Hygro, 100 ng of the AncYakut or Wrangel Island Mammoth NKD1 expression vector, and 10 ng of pRL-null with 0.1µL of PLUS reagent (Invitrogen) and 0.3 µL of Lipofectamine LTX (Invitogen) in 20 µL of

Opti-MEM. The cells were incubated in the transfection mixture for 6 hr until the transfection media was replaced with culture media and supplemented with the small molecule Wnt-signalig agonist CHIR99021. 48 hours after transfection, Dual Luciferase Reporter Assays (Promega) began with incubating the cells for 15 min in 20 µL of 1x passive lysis buffer. Luciferase and Renilla activity was measured with the Glomax multi+detection system (Promega). We standardized luciferase activity values to Renilla activity values and background activity values by measuring luminescence in wells lacking the NKD1 expression vector.

NEUROG3 functional validation

To infer if the G195E substitution had functional affects, we synthesized the ancestral mammoth (AncYakut, G195) and Wrangel Island (E195) *NEUROG3* genes with mouse codon usage and cloned each gene into the mammalian expression vector pcDNA3.1+C-DYK. Next we tested their ability to transactivate luciferase expression from the pGL3 luciferase reporter vector containing a minimal promoter and six repeats of the *PAX4* E-box (pGL3 [*luc/6x-PAX4E/minP*]); the *PAX4* E-box has previously been shown to physically bind NEUROG3 and drive luciferase expression in reporter assays [23]. African bush elephant primary dermal fibroblasts (San Diego Zoo, “Frozen Zoo”) were grown at 37°C/5% CO2 in a culture medium consisting of FGM/MEM (1:1) supplemented with 10% FBS, 1% Glutamine, and 1% penstrep. Confluent cells in 96-well plates in 60 µL of Opti-MEM (GIBCO) were transfected with 100 ng of the luciferase reporter plasmid 6x-PAX4E_pGL3-Basic, 100 ng of the AncYakut or Wrangel Island Mammoth NEUROG3 expression vector, and 10 ng of pRL-null with 0.1µL of PLUS reagent (Invitrogen) and 0.3 µL of Lipofectamine LTX (Invitogen) in 20 µL of Opti-MEM. The cells were incubated in the transfection mixture for 6 hr until the transfection media was replaced with culture media. 48 hours after transfection, Dual Luciferase Reporter Assays (Promega) began with incubating the cells for 15 min in 20 µL of 1x passive lysis buffer. Luciferase and Renilla activity was measured with the Glomax multi+detection system (Promega). We standardized luciferase activity values to Renilla activity values and background activity values by measuring luminescence in wells lacking the NEUROG3 expression vector.

REFERENCES

1. Palkopoulou, E. *et al.* Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25,** 1395–1400 (2015).
2. Palkopoulou, E., Dalén, L. & Lister, A. M. Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc R Soc B* (2013).
3. Nyström, V., Humphrey, J. & Skoglund, P. Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Mol. Ecol.* (2012).
4. Nyström, V. *et al.* Temporal genetic change in the last remaining population of woolly mammoth. *Proc. Biol. Sci.* **277,** 2331–2337 (2010).
5. Graham, R. W. *et al.* Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 9310–9314 (2016).
6. Rogers, R. L. & Slatkin, M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet.* **13,** e1006601 (2017).
7. Thomas, M. G. The flickering genes of the last mammoths. *Mol. Ecol.* **21,** 3379–3381 (2012).
8. Pečnerová, P., Díez-Del-Molino, D., Vartanyan, S. & Dalén, L. Changes in variation at the MHC class II DQA locus during the final demise of the woolly mammoth. *Sci Rep* **6,** 25274 (2016).
9. Dikov, N. N. The earliest sea mammal hunters of Wrangell Island. *Arctic Anthropology* (1988).
10. Gilbert, M. T. P. *et al.* Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 8327–8332 (2008).
11. Gilbert, M. T. P. *et al.* Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317,** 1927–1930 (2007).
12. Miller, W. *et al.* Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456,** 387–390 (2008).
13. Lynch, V. J. *et al.* Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep* **12,** 217–228 (2015).
14. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7,** Unit7.20 (2013).
15. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
16. Dammermann, A. *et al.* The hydrolethalus syndrome protein HYLS-1 links core centriole structure to cilia formation. *Genes Dev.* **23,** 2046–2059 (2009).
17. Badano, J. L., Mitsuma, N., Beales, P. L. & Katsanis, N. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* **7,** 125–148 (2006).
18. Angonin, D. & Van Raay, T. J. Nkd1 functions as a passive antagonist of Wnt signaling. *PLoS ONE* **8,** e74666 (2013).
19. Van Raay, T. J. *et al.* Naked1 antagonizes Wnt signaling by preventing nuclear accumulation of β-catenin. *PLoS ONE* **6,** e18650 (2011).
20. Van Raay, T. J., Coffey, R. J. & Solnica-Krezel, L. Zebrafish Naked1 and Naked2 antagonize both canonical and non-canonical Wnt signaling. *Dev. Biol.* **309,** 151–168 (2007).
21. Zhang, S. *et al.* Viable mice with compound mutations in the Wnt/Dvl pathway antagonists nkd1 and nkd2. *Mol. Cell. Biol.* **27,** 4454–4464 (2007).
22. Li, Q., Ishikawa, T.-O., Miyoshi, H., Oshima, M. & Taketo, M. M. A targeted mutation of

Nkd1 impairs mouse spermatogenesis. *J. Biol. Chem.* **280,** 2831–2839 (2005).
23. Smith, S. B., Watada, H. & German, M. S. Neurogenin3 activates the islet differentiation program while repressing its own expression. *Mol. Endocrinol.* **18,** 142–149 (2004).
24. Wang, J. *et al.* Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N Engl J Med* **355,** 270–280 (2006).
25. Pinney, S. E. *et al.* Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. *J. Clin. Endocrinol. Metab.* **96,** 1960–1965 (2011).
26. Rubio-Cabezas, O. *et al.* Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3. *Diabetes* **60,** 1349–1353 (2011).
27. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28,** 495–501 (2010).

AUTHOR CONTRIBUTIONS: E.F. performed NEUROG3 functional experiments and co-wrote the manuscript, S.K.K performed HYLS1 functional experiments and co-wrote the manuscript, S.C. and A.D performed NKD1 functional experiments, K. M. M. curated variants, A.R. performed genome analyses, B. J. M. performed NKD1 functional experiments and co-wrote the manuscript, W. M. performed genome analyses and co-wrote the manuscript, V. J. L. performed variant analyses and co-wrote the manuscript.

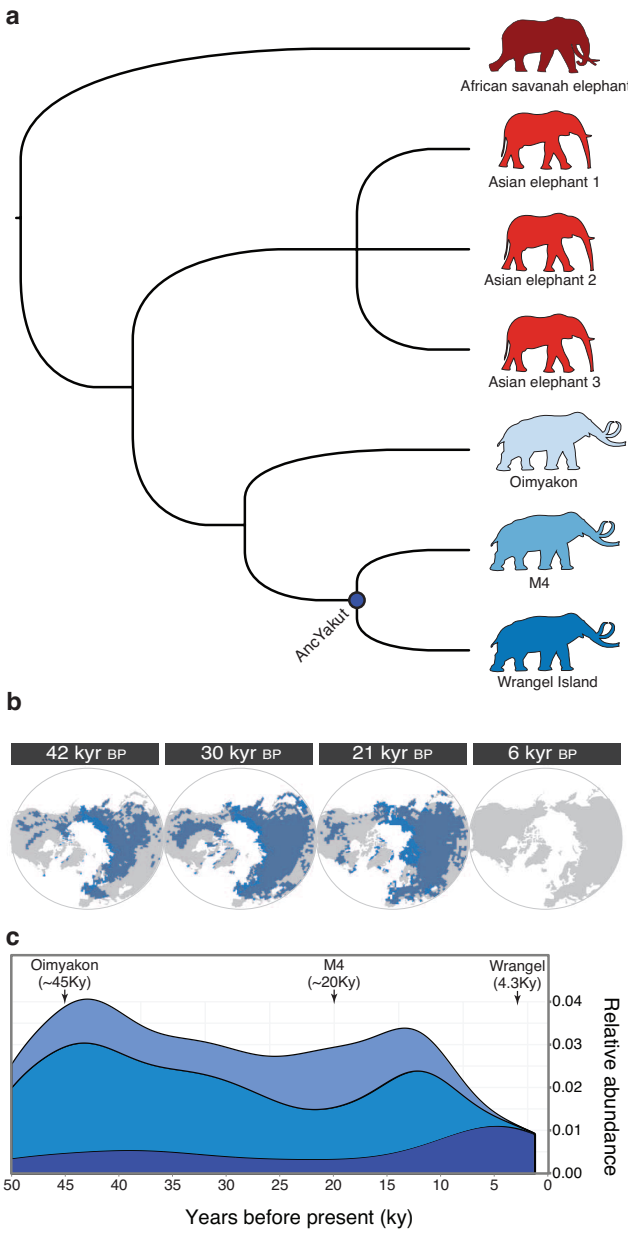

Figure 1. Demographic decline and extinction of Woolly Mammoths. (**a**) Phylogenetic relationship among African elephants, Asian elephants, and Woolly Mammoths. AncYakut, indicates node for which ancestral sequences were resurrected. (**b**) Range of Woolly Mammoths (blue) over the last 42,000 years. (**c**) Relative abundance of Woolly Mammoths over the last 50,000 years. Blue-gray, all populations. Blue, mainland populations. Dark blue, Island populations.

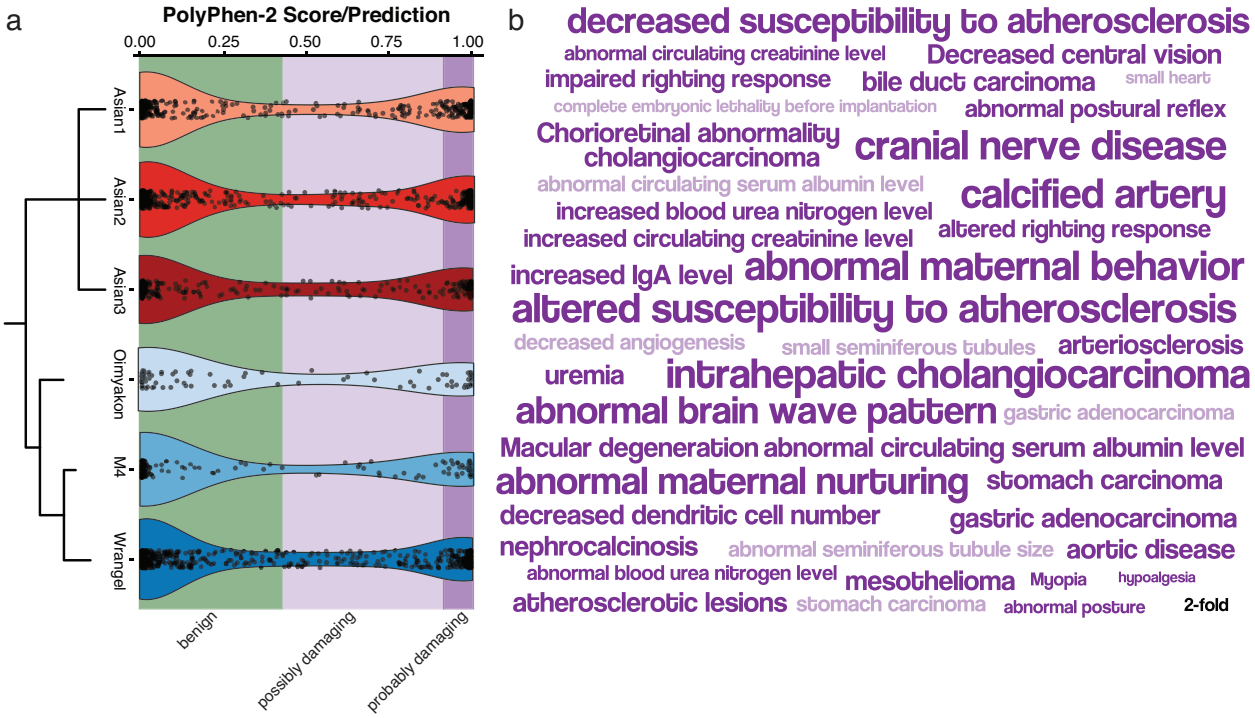

Figure 2. Accumulation and functional architecture of deleterious variants in the Wrangel Island mammoth genome. (**a**) Violin plot of PolyPhen-2 scores and predicted functional consequence (benign, possibly damaging, and probably damaging) for derived, homozygous variants in each individual. (**b**) Mouse KO phenotypes, human phenotypes, and disease ontologies in which fixed probably (dark purple) and possibly (light purple) damaging amino acid substitutions in the Wrangel Island mammoth are enriched. 2-fold inset, shows word size for 2-fold phenotypes and ontology enrichment.

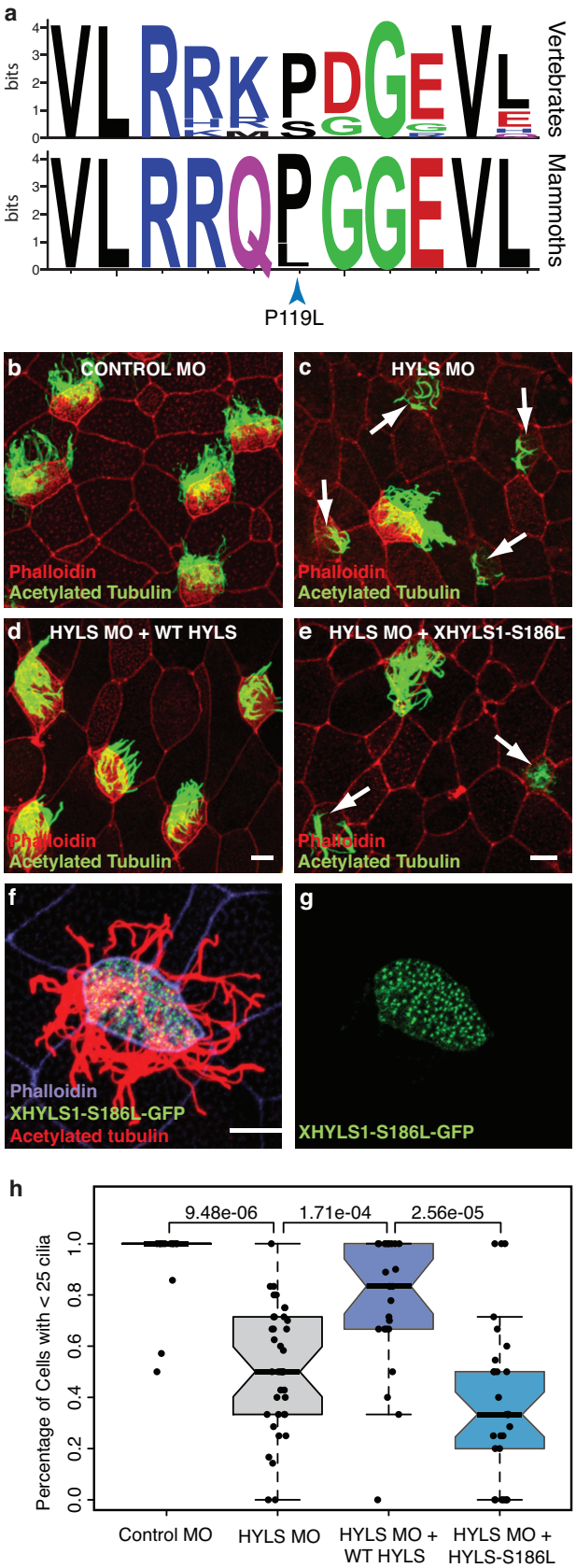

Figure 3. The Wrangel Island mammoth HYLS1 P119L mutant is dysfunctional. (**a**) Sequence logo showing conservation of HYLS1 around the Wrangel Island mammoth P119L mutation. Upper, logo from 100 vertebrates. Lower, logo from mammoths (Oimyakon, M4, and Wrangel Island). (**b-e**) Representative images of *Xenopus* multiciliated cells from embryos injected with Control MO (**b**), HYLS1 MO (**c**), HYLS1 MO and WT XHYLS1 (**d**) and HYLS1 MO and XHYLS1-S186L (**e**) stained with acetylated tubulin to mark cilia (green) and phalloidin to mark cell boundaries (red), note: arrows identify cells with defective ciliogenesis (scale bars = 10μm). (**f-g**) Multiciliated cell from a *Xenopus* embryo injected with mRNA encoding XHYLS1-S186L-GFP shows proper localization to the centrioles (green) and proper formation of cilia marked with acetylated tubulin (red) with cell boundaries marked with phalloidin (purple)(scale bar = 10μm). (**h**) Quantification of ciliogenesis defect scoring cells with < 25 cilia (Wilcox test *P*-values are shown).

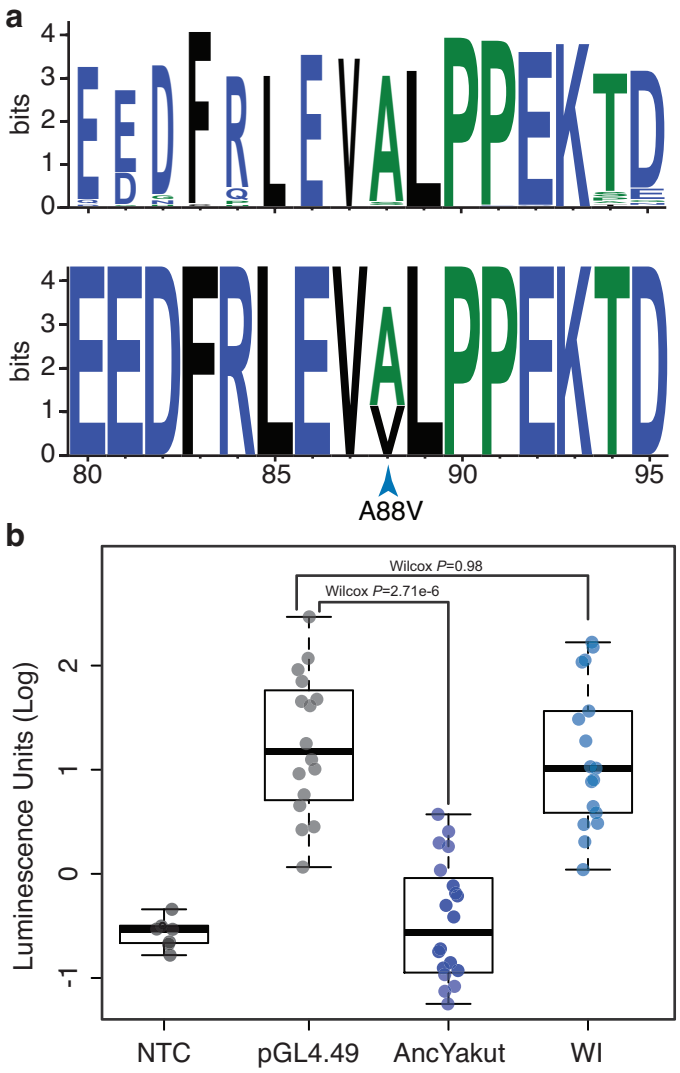

Figure 4. The Wrangel Island mammoth NKD1 A88V substitution is a loss of function mutation. (**a**) Sequence logo showing conservation of NKD1 around the Wrangel Island A88V mutation. Upper, logo from 100 vertebrates. Lower, logo from mammoths (Oimyakon, M4, and Wrangel Island). (**b**) Luciferase expression from the pGL4.49[*luc2P*/TCF-LEF/Hygro reporter vector in elephant fibroblasts transfected with control vector, AncYakut *NKD1*, or Wrangel Island (WI) Mammoth *NKD1*, and treated with the WNT signaling agonist CHIR99021. Background luminescence of non-transfected cells (NTC).

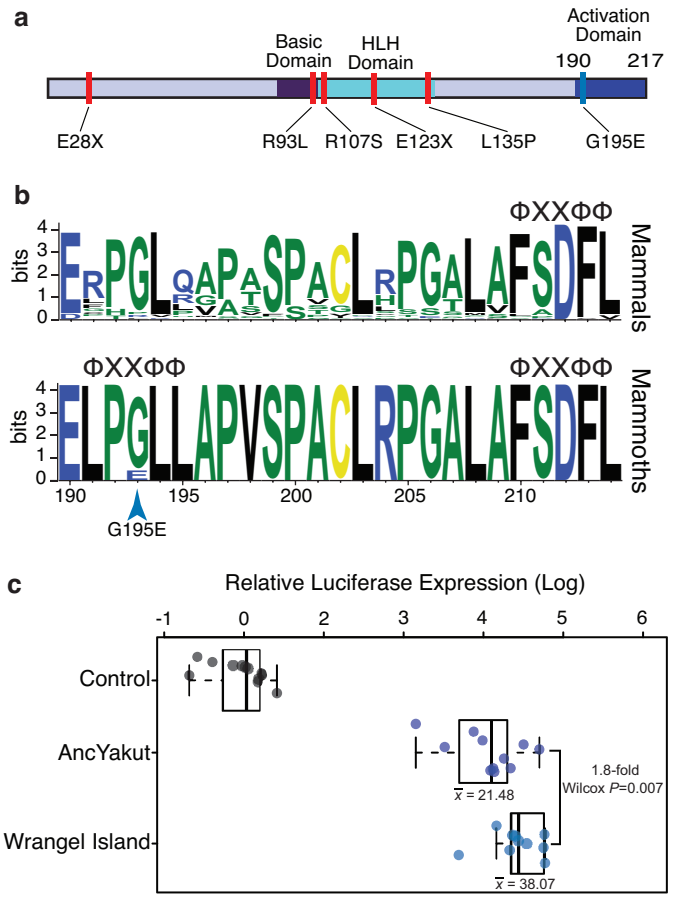

Figure 5. The Wrangel Island mammoth NEUROG3 is a hypermorph. (**a**) The Wrangel Island mammoth G195E mutation is located in the C-terminal transactivation domain of NEUROG3, location of mutations associated with human disease are shown in red. (**b**) Sequence logo showing conservation of NEUROG3 transactivation domain in mammals (upper) and mammoths (lower; Oimyakon, M4, and Wrangel Island). The location of ΦXXΦΦ co-factor interaction motifs are shown (Φ, any hydrophobic amino acid; X, any amino acid). (**c**) Relative luciferase expression from the pGL3 [*luc/6x-PAX4E/minP*] reporter vector in elephant fibroblasts transfected with control vector, AncYakut *NEUROG3*, or Wrangel Island Mammoth *NEUROG3*.

Table 1. PolyPhen-2 classification of private variants in each mammoth

Mammoth	probably damaging	possibly damaging	benign	unclassifiable
Oimyakon	19	18	65	4
M4	32	22	103	4
Wrangel Island	115	109	349	21